

## ORIGINAL ARTICLE

Ya Ping Qian · Zheng-Tao Chu · Qin Dai  
Can-Dong Wei · Jia You Chu · Atsushi Tajima  
Satoshi Horai

## Mitochondrial DNA polymorphisms in Yunnan nationalities in China

Received: December 22, 2000 / Accepted: January 18, 2001

**Abstract** Nucleotide sequences of the D-loop region of human mitochondrial DNA from four Yunnan nationalities, Dai, Wa, Lahu, and Tibetan, were analyzed. Based on a comparison of 563-bp sequences in 99 people, 66 different sequence types were observed. Of these, 64 were unique to their respective populations, whereas only 2 types were shared between the Lahu and Wa nationalities. The D-loop sequence variation and phylogenetic analysis suggested that the 99 mtDNA lineages were classified into eight clusters in the phylogenetic tree. All lineages that had a 9-bp deletion in the COII/tRNA<sup>Lys</sup> intergenic region appeared in one cluster in the D-loop tree, suggesting a single event of the deletion in the Yunnan nationalities studied. Genetic distances, based on net nucleotide diversities between populations including Han Chinese and mainland Japanese, revealed that the Dai, Wa, Lahu, and Han Chinese are closely related to each other, while Tibetan and mainland Japanese formed a single cluster. The bootstrap probability of separation between the Dai-Wa-Lahu-Chinese clade and the Tibetan-Japanese clade was 99%, indicating that there are at least two different origins among minority groups in Yunnan province. Although the genetic distance between Tibetan and Japanese within the clade is rather long, the results may shed light on the origins of mainland Japanese.

**Key words** Mitochondrial DNA · D-loop region · Sequence polymorphism · Intergenic COII/tRNA<sup>Lys</sup> 9-base pair deletion · Yunnan nationalities · Nucleotide diversity · Phylogenetic tree

Y.P. Qian · Z.-T. Chu · Q. Dai · C.-D. Wei · J.Y. Chu  
Institute of Medical Biology, Chinese Academy of Medical Sciences  
and Peking Union Medical College, Kunming, China

A. Tajima · S. Horai (✉)  
Department of Biosystems Science, The Graduate University for  
Advanced Studies, Hayama, Kanagawa 240-0193, Japan  
Tel. +81-468-58-1575; Fax +81-468-58-1544  
e-mail: horai@soken.ac.jp

### Introduction

Situated in southwest China, Yunnan is a multinational border province and is regarded as a hotspot for the study of the origin of Asian populations. In addition to the Han Chinese, there are 25 other ethnic groups living in the province, the smallest of which has a population of more than 4000. The population of these 25 ethnic minorities accounts for more than one-third of the total of 40 million minority people who live in the whole province. The various minority nationalities, consisting of 13 million people, are spread throughout the province, living peacefully in harmony and working together with the Han to promote common prosperity. On the basis of their historical records, there is some agreement that those populations speaking Sino-Tibetan are descendants of the Di-Qiang, those speaking Daic are descendants of the Baiyue, and the Austro-Asiatic speaking peoples are descendants of the Mon-Khmer. At present, Austro-Asiatic-speaking people live in southern Yunnan; Daic-speaking people live in southern, southeastern, and southwestern Yunnan; and most of the descendants of the Di-Qiang inhabit the central, northern, and western region of Yunnan province (You 1994). While it is true that some geographic barriers maintain genetic differences among such ethnic groups and provide valuable conditions in which to study their genetic relationship, genetic evidence has become necessary to verify such historical claims. Thus, the understanding of the origins and genetic relationship between each of these populations may shed light on the peopling of China, as well as that of other parts of Asia.

By virtue of its rapid rate of evolution and maternal mode of inheritance, mitochondrial DNA (mtDNA) has been widely used to study the genetic relationship among closely related populations (Horai et al. 1993; Horai et al. 1996). In recent mtDNA analyses, it has become common to directly sequence the major noncoding (D-loop) region of mtDNA that has apparently evolved several times faster than other parts (Vigilant et al. 1989; Horai and Hayasaka 1990; Horai et al. 1993; Horai et al. 1996). Sequence analysis

of this region therefore offers high resolution that discriminates among even very closely related individuals.

To investigate the genetic relationship and origins of Yunnan populations and to verify the historical records, we examined the mtDNA sequence variation in the D-loop region and intergenic COII/tRNA<sup>Lys</sup> 9-bp deletion in four Yunnan minority nationalities: Dai, Wa, Lahu, and Tibetan. It is believed, from their history, that the Tibetan and Lahu originate from the Di-Qiang tribe and that the Dai and Wa originate from the Baiyue and Mon-Khmer tribes, respectively. The results of the present study were analyzed together with the findings for large populations from China and Japan.

## Subjects and methods

### Subjects

Blood samples were obtained from 99 healthy individuals of four native Yunnan ethnic groups in China. Twenty-one Dai samples were collected from Xishuangbanna Prefecture, located in southwestern Yunnan; 22 Wa and 32 Lahu samples were collected from Simao Prefecture, in the southwestern part of this province; and 24 Tibetan samples were collected from Diqing Prefecture, which borders on Tibet. Genomic DNA was extracted either directly from lymphocytes or immortalized cell lines transformed by Epstein-Barr (EB) virus in our laboratory, by treatment with sodium dodecyl sulfate (SDS), proteinase K, and subsequent phenol/chloroform extraction. All the individuals gave their informed consent prior to their inclusion in this study.

### Direct sequencing of polymerase chain reaction (PCR) products from the D-loop region of mtDNA

A fragment of mtDNA was amplified by the PCR method described by Saiki et al. (1988). A set of two primers (A and E primers; see below) was designed to amplify a DNA fragment from the D-loop region of mtDNA (Horai et al. 1996). PCR was carried out under the following thermal cycle conditions: incubation at 94°C for 15s, 45°C for 15s, 72°C for 30s, for a total of 30 cycles. The amplified fragments were subjected to direct DNA sequencing, using the ABI PRISM Dye Terminator Cycle Sequencing FS Kit and ABI PRISM 377 DNA sequencer (Applied Biosystems, Foster City, CA, USA). The following oligonucleotide primers were designed for amplification and sequencing:

- A: 15897-5'-GTATAAACTAATACACCAGTCTTGT-3'-15921;
- B: 15985-5'-AGCACCCAAAGCTAAGATTC-3'-16004;
- C: 16204-5'-AGCAAGTACAGCAATCAACC-3'-16223;
- D: 16413-5'-GCGAGGAGAGTAGCACTCTT-3'-16432;
- E: 100-5'-CAGCGTCTCGCAATGCTATCGCGTG-3'-76.

The notation of Anderson et al. (1981) was used for the numbering of bases.

### Detection of the COII/tRNA<sup>Lys</sup> intergenic deletion

A set of oligonucleotide primers was designed, as follows, to amplify the COII/tRNA<sup>Lys</sup> intergenic region of mtDNA: 8211-5'-TCGTCCTAGAATTAATTCCC-3'-8230 and 8310-5'-AGTTAGCTTTACAGTGGGCT-3'-8291. This set of primers amplifies 100-bp and 91-bp fragments of mtDNA with two and one copies of the 9-bp tandem repeat (CCCCCTCTA), respectively. The amplified fragments were separated by electrophoresis on 4% NuSieve agarose (FMC BioProducts, Rockland, ME, USA) gels and detected fluorographically after staining with ethidium bromide.

### Data analyses

The number of nucleotide substitutions per site between individual sequences was estimated using the two-parameter model of nucleotide substitutions (Kimura 1980). On the basis of the estimated number of nucleotide substitutions, phylogenetic trees were constructed, using the neighbor-joining (NJ) method (Saitou and Nei 1987).

## Results

### Sequence differences and identities in Yunnan nationalities

We determined the nucleotide sequence of a 563-bp fragment of the D-loop region (positions 16048–16569, followed by positions 1–41 in the reference sequence of Anderson et al. 1981) for 99 people from the four Yunnan nationalities. There were 66 distinct types of sequences defined by 103 polymorphic sites (Fig. 1). Of these, 64 were unique to their respective populations, whereas only 2 (types 25 and 29) were shared, by five Lahu individuals and one Wa individual. Among the 64 unique types, 13 types (20%) were shared by more than two individuals within each population, whereas the remaining 51 types (80%) were observed in a single individual. The number of shared sequence types and the number of unique types observed in the four Yunnan populations are shown in Table 1.

All the sequence types observed in the Dai and Tibetan people were completely unique (100%), suggesting that these two ethnic groups are still maintaining their own ethnicity, with a lesser degree of genetic influence from other ethnic groups. By contrast, the Lahu and Wa have experienced some extent of gene flow between the two populations.

### Phylogeny of mtDNA sequences

The 563-bp sequences from the total of 99 individuals from Yunnan were aligned and compared, and the number of

**Fig. 1.** Nucleotide sequence differences in the D-loop region of mtDNA for 99 Yunnan people. Presented are nucleotides at 103 polymorphic sites based on a comparison of a 563-bp fragment of the D-loop region of mtDNA in the reference sequence of Anderson et al. (1981), shown below the other sequences, only the differences from the reference sequence are indicated. The number of individuals for each sequence type from the four populations investigated is shown with *Arabic numerals* at the right of each sequence. Eight clusters, characterized in Fig. 2, are shown with *brackets* with cluster names on the right

TYPE	Dai	Lahu	Wa	Tibetan
1	1	5		
2	1			
3	1			
4	1			
5	1			
6	1			
7	1			
8	1			
9	1			
10	1			
11	1			
12	1			
13	1			
14	1			
15	1			
16	1			
17	1			
18	1			
19	1			
20	1			
21	1			
22	1			
23	1			
24	1			
25	1			
26	1			
27	1			
28	1			
29	1			
30	1			
31	1			
32	1			
33	1			
34	1			
35	1			
36	1			
37	1			
38	1			
39	1			
40	1			
41	1			
42	1			
43	1			
44	1			
45	1			
46	1			
47	1			
48	1			
49	1			
50	1			
51	1			
52	1			
53	1			
54	1			
55	1			
56	1			
57	1			
58	1			
59	1			
60	1			
61	1			
62	1			
63	1			
64	1			
65	1			
66	1			

**Table 1.** Unique and common sequence types observed in four Yunnan populations

Population	No. of individuals	No. of sequence types	No. of unique types (% of individuals) <sup>a</sup>	No. of common types shared between two populations
Dai	21	21	21 (100%)	0
Lahu	32	13	11 (68.8%)	2 (t25, t29) <sup>b</sup>
Tibetan	24	18	18 (100%)	0
Wa	22	16	14 (90.9%)	2 (t25, t29)
Total	99	66	64 (87.9%)	—

<sup>a</sup> Percentages of individuals who possess unique types in each population

<sup>b</sup> Shared types, in parenthesis, are indicated by “t” followed by the type number

**Table 2.** Shared polymorphic sites in the clusters

Cluster	No. of lineages	Shared polymorphic sites within a cluster <sup>a</sup>					
C1	13	<b>9-bp deletion (100%)</b>	16189:T/C (100%)	<b>16217:T/C (92%)</b>	<b>16261:C/T (54%)</b>	16519:T/C (100%)	
C2	5	<b>16145:G/A (100%)</b>	16148:C/T (80%)	<b>16188:C/T (80%)</b>	16189:T/C (100%)	16223:C/T (80%)	<b>16381:T/C (80%)</b>
C3	7	16192:C/T (43%)	<b>16304:T/C (100%)</b>	<b>16309:A/G (57%)</b>	<b>16390:G/A (86%)</b>	16519:T/C (71%)	
C4	23	16129:G/A (83%)	<b>16172:T/C (91%)</b>	<b>16304:T/C (100%)</b>	16519:T/C (100%)		
C5	7	16185:C/T (71%)	16223:C/T (100%)	<b>16260:C/T (86%)</b>	<b>16298:T/C (100%)</b>		
C6	8	<b>16182:A/G (50%)</b>	<b>16193:C/T (50%)</b>	16223:C/T (100%)	16519:T/C (100%)		
C7	18	16129:G/A (56%)	16223:C/T (100%)	<b>16290:C/T (61%)</b>	<b>16319:G/A (50%)</b>	16519:T/C (61%)	
C8	18	16223:C/T (100%)	16362:T/C (89%)				

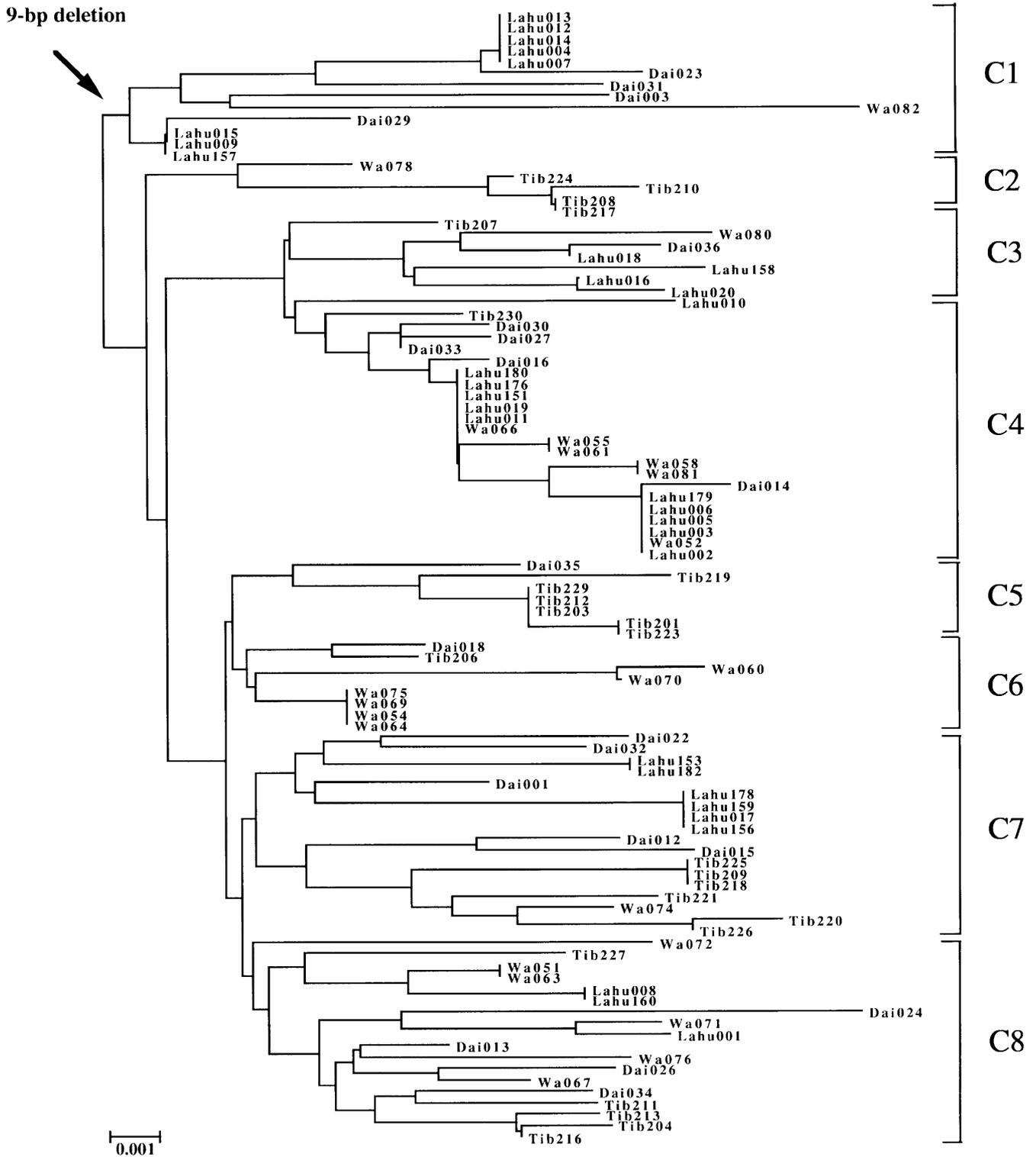
Numbers in parentheses represent the percentages of lineages (more than 40%) that exhibited the polymorphism. Polymorphic sites, shown in boldface, are specific to each cluster

<sup>a</sup> Each polymorphic site corresponds to the site and replacement in the reference sequence reported by Anderson et al. (1981)

nucleotide substitutions between each pair of sequences was estimated. Nucleotide diversity among the 99 individuals was estimated to be 1.48%, which is slightly higher than the corresponding value estimated for East Asian populations (1.34%; Horai et al. 1996). Using the estimated number of nucleotide substitutions between individual sequences, we constructed a phylogenetic tree, using the neighbor-joining (NJ) method (Saitou and Nei 1987), as shown in Fig. 2. On the basis of branching patterns in the tree, sequences were classified into eight monophyletic clusters, which are shown in Fig. 2 by brackets with cluster numbers (C1–C8). To ensure the robustness of these clusters, we examined shared polymorphic sites which appeared in more than 40% of lineages within each cluster, as shown in Table 2. The majority of members in each cluster exhibited a specific combination of two or three polymorphic sites, with the exception of members of C8, who do not have any specific polymorphic sites. Most members of clusters C2, and C5 through C8, shared a C-to-T transition at 16223 (nucleotide position in the reference sequence of Anderson et al. 1981), while this polymorphism was virtually absent in clusters C1, C3, and C4. A T-to-C transition at 16519 was the overwhelming polymorphism in the entire population, and this polymorphism occurred in seven clusters, except for C5. A G-to-A transition at 16129 occurred in the majority of lineages from C4 (83%) and C7 (56%), whereas these two clusters appeared at discrete positions in the phylogenetic tree. Thus, the above three polymorphisms are due to either recurrent mutations or ancient polymorphisms. A T-to-C transition at 16362 was predominant in C8 (89%), whereas this polymorphism was virtually absent in other

clusters, except for C7 (28%). Although this polymorphism appeared to have occurred in parallel during mitochondrial lineage divergence, it is informative for late-diverging lineages (C7 and C8) in the phylogenetic tree. On the other hand, in some clusters, most of the members exhibited a specific polymorphism that was virtually absent in other clusters. For example, a T-to-C transition at 16298 was observed in all the members (100%) of C5, which was unique to this cluster and was completely absent in other clusters. Moreover, in C3 and C4, all lineages exhibited a T-to-C transition at 16304, which is specific to these two clusters. Unique polymorphisms were also seen in four other clusters, as shown in Table 2. Therefore, it is likely that most of the clusters could reflect their ancestral states of lineage divergence. However, it is apparent that lineages from the four populations studied were intermingled in the phylogenetic tree, although individuals from single populations dominated in some of the clusters.

To evaluate the features of the cluster patterns further, the composition of the eight clusters and the ethnic origins of the 99 Yunnan lineages were summarized, as shown in Table 3. Although the numbers of individuals sampled from the four populations were not the same, we assigned “specificity” for each cluster based on the population from which the maximum percentage of individuals was derived. Although this assignment of specificity seems to be somewhat arbitrary, the rule for assignment is a simple “majority-rules” voting procedure. Furthermore, it may be useful for understanding the relationships of mtDNA sequences among very closely related human populations, such as Yunnan ethnic groups, because ancient migrations between



**Fig. 2.** Phylogenetic tree showing the 99 Yunnan mtDNA lineages from four populations. The phylogenetic tree was constructed, based on the pairwise number of nucleotide substitutions in the 563-bp D-loop region, by the neighbor-joining (NJ) method (Saitou and Nei 1987). The eight distinct clusters in the tree are indicated by brackets, with cluster numbers C1-C8. The names of the populations (Dai, Lahu,

Wa, and Tibetan) are indicated by the abbreviations at the tip of each branch: the three digits following the population names are the individual numbers of the sample. All the individuals with the 9-bp deletion are shown after the branching points indicated by the arrow. The number of nucleotide substitutions per site is indicated on the scale below the tree

**Table 3.** Distribution of the four ethnic groups in each monophyletic cluster of the phylogenetic tree

Cluster	Specificity	Dai	Lahu	Tibetan	Wa	Total
C1	Lahu-1	30.8	<b>61.5</b>	0	7.7	13
C2	Tibetan-1	0	0	<b>80.0</b>	20.0	5
C3	Lahu-2	14.3	<b>57.1</b>	14.3	14.3	7
C4	Lahu-3	21.7	<b>47.8</b>	4.3	26.1	23
C5	Tibetan-2	14.3	0	<b>85.7</b>	0	7
C6	Wa-1	12.5	0	12.5	<b>75.0</b>	8
C7	Unassigned-1	27.8	33.3	33.3	5.6	18
C8	Wa-2	22.2	16.7	27.8	<b>33.3</b>	18

Percent values for individuals in each cluster are given. The maximum value in the cluster is shown in boldface

populations are assumed. For example, in C1, Lahu were dominant (62%), with four members from the Dai and one from the Wa. Lahu was also dominant in C3 (57%) and C4 (48%). We therefore assigned these clusters as being specific to the Lahu and named them Lahu-1, Lahu-2, and Lahu-3 (Table 3). In this way, we were able to assign specificity for seven clusters, except for C7. Because C7 had one-third of its members from the Lahu and one-third from the Tibetan populations, we were unable to assign the specificity of this cluster. It is interesting to note that individuals from the Dai were not dominant in any clusters, although this population may have influenced gene flow toward the other three ethnic groups. On the other hand, Tibetans dominated in two clusters, C2 (80%) and C5 (86%), although these two clusters contained five and seven individuals, respectively. This suggests that some lineages from the Tibetan population exhibited unique phylogenetic positions among the four present ethnic groups.

#### COII/tRNA<sup>Lys</sup> intergenic 9-bp deletion

The presence of a 9-bp deletion in the COII/tRNA<sup>Lys</sup> intergenic region of mtDNA is one of the characteristics not only of Asians (Horai and Matsunaga 1986; Horai 1987; Stoneking and Wilson 1989; Horai 1991a, 1991b; Ballinger et al. 1992; Harihara et al. 1992; Passarino et al. 1993; Horai et al. 1996) but also of populations of Asian origin, including Polynesians (Hertzberg et al. 1989; Hagelberg and Clegg 1993; Lum et al. 1994; Redd et al. 1995) and Native Americans (Schurr et al. 1990; Ward et al. 1991; Shields et al. 1992, 1993; Torroni et al. 1992; Horai et al. 1993). However, some studies have suggested multiple origins for the deletion in Asia (Schurr et al. 1990; Ballinger et al. 1992; Barrientos et al. 1995; Redd et al. 1995). We screened the present Yunnan people from four localities for the 9-bp deletion. For each population, the frequency of the deletion was 25% for the Lahu, 19% for the Dai, and 5% for the Wa; however, in Tibetans, the 9-bp deletion was completely absent, indicating again the unique genetic feature of this ethnic group. The frequency of the 9-bp deletion was 12% in the entire sample and all of the individuals with this deletion appeared in C1 of the phylogenetic tree (Fig. 2), indicating a single origin of the deletion among the four Yunnan nationalities.

#### Variation in the hypervariable domain of the D-loop region

In the D-loop region, there is a hypervariable domain that encompasses from bp 16180 to bp 16193. Because of length differences between individuals, we excluded two sites (bp 16182, 16183) and the elongated C stretch from the present phylogenetic analysis. However, this domain is informative for making phylogenetic inferences between populations. There were 25 sequence types in the 14-bp hypervariable domain in eight populations (including the four populations reported by Horai et al. 1993; see Table 4); types 1–19 had been reported before, types 20–25 were new types found in the four Yunnan populations. As in the four populations reported previously, type 1 was the most frequent sequence type in the four Yunnan populations, with frequencies of 54%–69%; about 59% of all the individuals in the eight populations exhibited the sequence that contains four As, five Cs, one T, and four Cs in this order. Twelve variant sequences were due to base substitutions in the C stretch (types 2–10 and three new types, 20–22). In other variant sequences, differences resulted from elongation of the C stretch and shortening of the A stretch (types 11–19 and three new types, 23–25). These distinct sequence patterns in the hypervariable region indicate that once the T at bp 16189 is substituted by C, the number of As and Cs becomes flexible, probably as a result of errors in replication (Horai et al. 1993). Furthermore, all mtDNA lineages which had a 9-bp deletion in the COII/tRNA<sup>Lys</sup> intergenic region possessed the T-to-C transition and resulting flexible numbers of As and Cs (Fig. 1). In the present study, types 5–9, 11, 12, 15, and 16 were not found in the four Yunnan populations. It is also noted that a T-to-C transition at 16249, which was found only in a Native American (Horai et al. 1993), was exhibited in a Wa individual (type 5 in Fig. 1).

#### Nucleotide diversity and phylogeny in Yunnan populations

Nucleotide diversity is one of the important measures of inferring the demographic process in a population. Using intrapopulation nucleotide diversity ( $d_x$  or  $d_y$ ), and interpopulation nucleotide diversity ( $d_{xy}$ ), we can also estimate net nucleotide diversity ( $d_A$ ) between each pair of populations, and  $d_A$  can be used to infer relatedness among populations (Nei and Miller 1990; Horai et al. 1996). In this study, we estimated nucleotide diversity values for the four Yunnan populations, together with those of two large Han Chinese and mainland Japanese populations (Horai et al. 1996), as shown in Table 5. The intrapopulation nucleotide diversity values ( $d_x$  or  $d_y$ ) ranged from 1.346% to 1.689%. However, the net nucleotide diversity ( $d_A$ ) values between the Tibetan and the other Yunnan groups were high (ranging from 0.137% to 0.350%), indicating that the Tibetan population has a unique genetic position among Yunnan populations. Of note is that the lowest  $d_A$  value was observed between the Tibetan and Japanese populations (0.09%) in comparisons of Tibetan and other populations.

**Table 4.** mtDNA sequence types in a hypervariable domain, and observed numbers in four Yunnan and four other populations

Type	Sequence	Dai	Wa	Lahu	Tibet	Eu	Afr	As	Am	Number	Frequency
1	AAAA----CCCCCTCCCC	12	14	22	13	15	5	55	36	172	0.589
2	AAAA----CCCCCTCCCT		4			1		1		6	0.021
3	AAAA----CCCCCTCCTC	3		1				1	2	7	0.024
4	AAAA----CTCCCTCCCC				5			2		7	0.024
5	AAAA----CCCTCTCCCC							3	11	14	0.048
6	AAAA----CCCTCCCCCC						2	1	2	5	0.017
7	AAAA----CCCTGCCCCC						1			1	0.003
8	AAAA----TCCTACCCCC						1			1	0.003
9	AAAA----CCTCCCCCCC					1				1	0.003
10	AAAA----CCCCCCCCCC	1				1		7	4	13	0.045
11	AAAA---CCCCCCCCCCC					1		1	1	3	0.010
12	AAA-CCCCCCCCCCCCC							1		1	0.003
13	AAA--CCCCCCCCCCCCC	1						1	1	3	0.010
14	AAA---CCCCCCCCCCCC				1	1		8	8	18	0.062
15	AAA----CCCCCCCCCCCC								2	2	0.007
16	AAA-----CCCCCCCCC							1		1	0.003
17	AA--CCCCCCCCCCCCC		1	4				6		11	0.038
18	AA---CCCCCCCCCCCC	1		5			1	3		10	0.034
19	AA----CCCCCCCCCCCC	1							5	6	0.021
20	AAAA---CTCCCCCCCC				1					1	0.003
21	AAA---CACCCCCCCCC	1								1	0.003
22	AAAA--CCCCCCCCCCCC		2							2	0.007
23	AA---CCCCCTCCCCC	1								1	0.003
24	AAAA----CCCCTCCCC				4					4	0.014
25	AAAA----CCCCCTCCC		1							1	0.003
Total		21	22	32	24	20	10	91	72	292	1.000

Eu, Af, As and Am indicate European, African, Asian, and Native American populations, respectively; these data are from Horai et al. (1993)

**Table 5.** Estimates of interpopulational ( $d_{xy}$ ), intrapopulational ( $d_x$  or  $d_y$ ), and net ( $d_A$ ) nucleotide diversity among six populations

Population	Dai ( $n = 21$ )	Tibetan ( $n = 24$ )	Wa ( $n = 22$ )	Lahu ( $n = 32$ )	Han Chinese ( $n = 66$ )	Japanese ( $n = 62$ )
Dai	<b>1.689</b>	1.767	1.616	1.576	1.672	1.545
Tibetan	0.137	<b>1.571</b>	1.720	1.832	1.759	1.548
Wa	0.002	0.166	<b>1.537</b>	1.523	1.616	1.497
Lahu	0.033	0.350	0.057	<b>1.395</b>	1.638	1.575
Han Chinese	0	0.136	0.010	0.103	<b>1.676</b>	1.534
Japanese	0.027	0.090	0.055	0.205	0.022	<b>1.346</b>

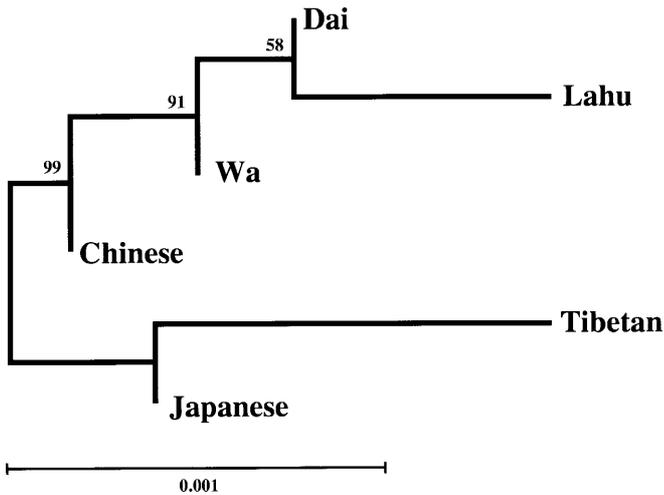
All values are multiplied by 100. The figures on the diagonal (shown in boldface) refer to ( $d_x$  or  $d_y$ ), and those above the diagonal refer to  $d_{xy}$ ; the figures below the diagonal represent the value,  $d_A = d_{xy} - [(d_x + d_y)/2]$

Low  $d_A$  values were observed between Dai and Han Chinese (0.00%), between Dai and Wa (0.002%), and between Wa and Han Chinese (0.010%), indicating close genetic relationships between these three groups of populations (Dai, Wa, and Han Chinese). By contrast,  $d_A$  values between the Lahu and three other populations (Dai, Wa, and Han Chinese) in China were intermediate (range, 0.033% to 0.103%), although  $d_A$  value between Lahu and Japanese was estimated at 0.205%. On the basis of  $d_A$  distances, a phylogenetic tree was constructed, using the NJ method (Fig. 3). The tree clearly indicated that the six populations listed in Table 5 were separated into two clades. The first clade included Han Chinese, Wa, Dai, and Lahu, while the second clade consisted of Japanese and Tibetan populations, although the length of the branch between these two populations was not so short. The separation between the first and second clades was significantly supported by a high bootstrap probability (99%). In this tree, the clustering of

three Yunnan nationalities (Wa, Dai, and Lahu) was also supported by a high bootstrap probability (91%), suggesting that the three groups shared some magnitude of maternal ancestry. Furthermore, the Wa and Dai appeared in close proximity in the tree, suggesting that the genetic distance between them was coincident with their geographic distribution.

## Discussion

Restriction enzyme analysis of mtDNA has been a powerful tool in attempts to quantify evolutionary relationships among human ethnic groups (Cann et al. 1987; Merriwether et al. 1991; Torroni et al. 1994a, 1994b, 1994c). Although restriction enzyme analyses represent polymorphism over the whole mitochondrial genome, some ambiguities remain with respect to the actual number of nucleotide differences



**Fig. 3.** Phylogenetic tree, constructed by the neighbor-joining (NJ) method, showing the relationships of six populations, based on net nucleotide diversity ( $d_A$ ) distances. The *number* for each interior branch is the bootstrap probability (Felsenstein 1986). The  $d_A$  distances are indicated on the *scale* below the tree

and with respect to the estimation of genetic distances (Kocher et al. 1989; Horai and Hayasaka 1990). However, sequence analysis of the mtDNA D-loop region directly offers reliable results and has been used in evolutionary studies (Vigilant et al. 1989, 1991; Soodyall et al. 1996; Horai et al. 1996; Bonatto and Salzano 1997; Starikovskaya et al. 1997).

In the present study, we analyzed the 563-bp sequence in the mtDNA D-loop region and the deletion in the COII/tRNA<sup>Lys</sup> intergenic region in 99 individuals of four Yunnan nationalities, Dai, Wa, Lahu, and Tibetan, to determine the origins of these populations and their genetic relationships with each other. Based on a comparison of 563-bp sequences in the 99 people, there were 66 different sequence types, defined by 103 polymorphic sites. Of these 66 sequence types, 64 were unique to their respective populations, whereas only 2 types were shared, between the Lahu and Wa nationalities. All the sequence types observed in the Dai and Tibetan populations were completely unique, suggesting that these ethnic groups are still maintaining their own ethnicity with a lesser degree of genetic influence from other ethnic groups.

As one of the characteristics of Asians, Native Americans, and some Pacific island populations (Hertzberg et al. 1989; Stoneking et al. 1990; Ballinger et al. 1992), the 9-bp deletion in the COII/tRNA<sup>Lys</sup> intergenic region was also detected in this study. For each population, the frequency of the deletion was 25% for the Lahu, 19% for the Dai, and 5% for the Wa; however, in Tibetans, the 9-bp deletion was completely absent, indicating the unique genetic feature of this ethnic group. Some studies have suggested multiple origins for the deletion in Asia (Schurr et al. 1990; Ballinger et al. 1992; Redd et al. 1995), although the deletion may have occurred only once during the evolution of modern types of human mtDNA (Wrischnik et al. 1987). Also, a comparison of the D-loop region sequence data from Asian and sub-Saharan African populations with the deletion in-

dicated that the 9-bp deletion arose independently in sub-Saharan Africa and Asia, and that the deletion has arisen more than once in Africa (Soodyall et al. 1996). The D-loop sequence variation and phylogenetic analysis in the current study suggested that the 99 mtDNA lineages were classified into eight clusters in the phylogenetic tree. All lineages that had the 9-bp deletion in the COII/tRNA<sup>Lys</sup> intergenic region appeared in one cluster in the D-loop tree, suggesting a single event of the deletion in the Yunnan nationalities studied. Furthermore, we compared the hypervariable domain sequences of the D-loop region from deletion-positive individuals in the four ethnic groups in this study with those in European, Asian, Native American, and African populations (Horai et al. 1993; Soodyall et al. 1996), indicating the certain Asian origin of the deletion observed in the four Yunnan populations.

In the historical record, the Dai are considered to be a population that originated from the coastal area of south-eastern China thousands of years ago. On the other hand, there were some Mon-Khmer tribes living in Yunnan province a long time ago, but most of them emigrated to south-eastern Asia around 2000 BC. The Wa are classified as one of a few such populations that settled in Yunnan province (You 1994). It is believed that the Lahu and Tibetan populations originated from the Di-Qiang tribe, and these two populations speak similar languages that belong to the Tibetan-Burma language subfamily. Without large-scale migration and hybridization after settlement in a local area, especially for the Lahu, Wa, and Tibetan populations, these populations may have lived in isolated areas with great geographic boundaries for thousands of years. In addition, matrilineal clanship and a style of paramount maternal power in the clan have continued for a long time in the history of the Lahu (Shi 1997). By contrast, the Dai live scattered over a vast area of the southwest part of Yunnan province, with a population of more than one million. On the basis of net nucleotide diversities between populations, the Dai, Lahu, and Wa converged very closely together, leaving the Tibetan population in the NJ tree with a high bootstrap probability. This observation also suggests the uniqueness of the Tibetan population among the four ethnic groups, although it does not entirely support the historical records or linguistic evidence. Furthermore, coordinated research on the genetic relationships in 28 Chinese ethnic groups, carried out by the genotyping of microsatellite polymorphic loci, revealed that Chinese populations diverged into northern and southern groups (Chu et al. 1998). The Lahu were clearly shown to be in the southern cluster with the Dai and Aini and others in the phylogenetic tree. Also, the Tibetan population was shown to be in the northern cluster in that tree (Chu et al. 1998). Although historical studies may not shed light on the true relationship, the finding of close genetic affinity among the Dai, Wa, and Lahu indicates that genetic hybridization may have occurred, assuming that the historical records of their origins are correct. Our recent study of the origins of the Tibetan population, carried out by genotyping some sets of Y-chromosome single-nucleotide polymorphic loci, showed the multiple origins of the Tibetan population; one was

from central Asia and another from East Asia (Qian et al. 2000). This result suggested that one interpretation could be used to explain the great genetic differences between the Lahu and Tibetan populations.

The present study also shed light on the origins of the Japanese, because mainland Japanese and Tibetan populations formed a single cluster, separating from the Wa, Dai, and Lahu from Yunnan, as well as Han Chinese, with a very high bootstrap probability (99%). This notion coincided with the previous observation of Y-chromosome polymorphisms (Hammer and Horai 1995; Karafet et al. 1999), where the frequency of the *Alu* insertion polymorphism (YAP+) was very high in Tibetan, as well as in several Japanese populations.

In this study, we focused on the four Yunnan ethnic populations to infer genetic relationships among these people. Further study will be necessary for many of the remaining Yunnan nationalities to reconstruct a more realistic scenario of the peopling of Yunnan province, as well as that of Asia.

**Acknowledgments** This work was supported by a grant from the National Science Foundation of China, a grant from the Chinese Medicine Board, USA (J.Y. Chu), and grant to young researchers from the Science and Technology Foundation of Yunnan province (Y.P. Qian) and grants from the Ministry of Education, Science, and Culture of Japan (S. Horai).

## References

- Anderson S, Bankier AT, Barrell BG, Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465
- Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen KH, Wallace DC (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migration. *Genetics* 130:139–152
- Barrientos A, Casademont J, Solans A, Moral P, Cardellach F, Urbano-Marquez A, Estivill X, Nunes V (1995) The 9-bp deletion in region V of mitochondrial DNA: evidence of mutation recurrence. *Hum Genet* 96:225–228
- Bonatto SL, Salzano FM (1997) Diversity and age of the four major mtDNA haplogroups and their implications for the peopling of the new world. *Am J Hum Genet* 61:1413–1423
- Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325:31–36
- Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ, Chu ZT, Yang ZQ, Lin KQ, Li P, Wu M, Geng ZC, Tan CC, Du RF, Jin L (1998) Genetic relationship of populations in China. *Proc Natl Acad Sci USA* 95:11763–11768
- Felsenstein J (1986) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791
- Hagelberg E, Clegg JB (1993) Genetic polymorphisms in prehistoric Pacific islanders determined by analysis of ancient bone DNA. *Proc R Soc Lond [Biol]* 252:163–170
- Hammer MF, Horai S (1995) Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet* 56:951–962
- Harihara S, Hirai M, Suutou Y, Shimizu K, Omoto K (1992) Frequency of a 9-bp deletion in the mitochondrial DNA among Asian populations. *Hum Biol* 66:161–166
- Hertzberg M, Mickleleson KNP, Serjeantson SW, Prior JF, Trent RJ (1989) An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. *Am J Hum Genet* 44: 504–510
- Horai S (1987) Evolutionary implications of mitochondrial DNA polymorphism in human populations. In: Vogel F, Sperling K (eds) *Human genetics, proceedings of the 7th international congress*. Springer-Verlag, Heidelberg Berlin Tokyo New York, pp 177–181
- Horai S (1991a) Molecular phylogeny and evolution of human mitochondrial DNA. In: Kimura M, Takahata N (eds) *New aspects of the genetics in molecular evolution*. Japan Scientific Societies Press, Tokyo/Springer-Verlag, Berlin Heidelberg Tokyo New York, pp 135–152
- Horai S (1991b) A genetic trail of human mitochondrial DNA. In: Mukohata Y (ed) *New era of bioenergetics*. Academic, Tokyo, pp 273–299
- Horai S, Hayasaka K (1990) Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am J Hum Genet* 46:828–842
- Horai S, Matsunaga E (1986) Mitochondrial DNA polymorphism in Japanese. II. Analysis with restriction enzymes of four or five base pair recognition. *Hum Genet* 72:105–117
- Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonoda S, Tajima K (1993) Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Mol Biol Evol* 10:23–47
- Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, Park KS, Omoto K, Pan IH (1996) mtDNA polymorphism in East Asian population, with special reference to the peopling of Japan. *Am J Hum Genet* 59:579–590
- Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF (1999) Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am J Hum Genet* 64:817–831
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120
- Kocher TD, Thomas WK, Meyer A, Edwards SV, Paabo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc Natl Acad Sci USA* 86:6196–6200
- Lum JK, Rickards O, Ching C, Cann RL (1994) Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. *Hum Biol* 66:567–590
- Merriwether DA, Clark AG, Ballinger SW, Schurr TG, Soodyall H, Jenkins T, Sherry ST, Wallace DC (1991) The structure of human mitochondrial DNA variation. *J Mol Evol* 33:543–555
- Nei M, Miller JC (1990) A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. *Genetics* 125:873–879
- Passarino G, Semino O, Modiano G, Santachiara-Benerecetti AS (1993) COII/tRNA<sup>Lys</sup> intergenic 9-bp deletion and other mtDNA markers clearly reveal that the Tharus (southern Nepal) have Oriental affinities. *Am J Hum Genet* 53:609–618
- Qian YP, Qian BZ, Su B, Yu JK, Ke YH, Chu ZT, Shi L, Chu JY, Lu DR, Jin L (2000) Multiple origins of Tibetan Y chromosomes. *Hum Genet* 106:453–454
- Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro ASM, Stoneking M (1995) Evolutionary history of the COII/tRNA<sup>Lys</sup> intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol Biol Evol* 12:604–615
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi RG, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487–491
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, Wallace DC (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am J Hum Genet* 46:613–623
- Shi J (1997) Situation of society and style of economy and culture in Southwestern Chinese nationalities. Yunnan Education Press, Kunming, pp 1–8
- Shields GF, Hecker K, Voevoda MI, Reed JK (1992) Absence of the Asian-specific region V mitochondrial marker in Native Beringians. *Am J Hum Genet* 50:758–765
- Shields GF, Schmiechen AM, Frazier BL, Redd A, Voevoda MI, Reed JK, Ward RH (1993) MtDNA sequences suggest a recent evolution-

- ary divergence for Beringian and Northern North American populations. *Am J Hum Genet* 53:549–562
- Soodyall H, Vigilant L, Hill AV, Stoneking M, Jenkins T (1996) mtDNA control-region sequence variation suggests multiple independent origins of an “Asian-specific” 9-bp deletion in sub-Saharan Africans. *Am J Hum Genet* 58:595–608
- Starikovskaya YB, Sukernik RI, Schurr TG, Kogelnik AM, Wallace DC (1998) mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of ancient Beringia and the peopling of the New World. *Am J Hum Genet* 63:1413–1423
- Stoneking M, Wilson AC (1989) Mitochondrial DNA. In: Hill A, Serjeantson S (eds) *The colonization of the Pacific: a genetic trail*. Clarendon, Oxford, pp 215–245
- Stoneking M, Jord LB, Bhatia K, Wilson AC (1990) Geographic variation on human mitochondrial DNA from Papua New Guinea. *Genetics* 124:717–733
- Torroni A, Schurr TG, Yang C, Szathmary EJE, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, Wallace DC (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130:153–162
- Torroni A, Chen Y, Semino O, Santachiara-Beneceretti AS, Scott CR, Lott MT, Winter M, Wallace DC (1994a) mtDNA and Y-chromosome polymorphisms in four Native American populations from Southern Mexico. *Am J Hum Genet* 54:303–318
- Torroni A, Lott MT, Cabell M, Chen YS, Lavergne L, Wallace DC (1994b) mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-Loop region. *Am J Hum Genet* 55:760–776
- Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, Wallace DC (1994c) Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. *Am J Phys Anthropol* 93:189–199
- Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC (1989) Mitochondrial DNA sequences in single hairs from a southern African population. *Proc Natl Acad Sci USA* 86:9350–9354
- Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. *Science* 253:1503–1507
- Ward RH, Frazier BL, Dew-Jager K, Paabo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. *Proc Natl Acad Sci USA* 88:8720–8724
- Wrischnik LA, Higuchi RG, Stoneking M, Erlich HA, Arnheim N, Wilson AC (1987) Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. *Nucleic Acids Res* 15:529–542
- You Z (1994) *History of Yunnan nationalities*. Yunnan University Press, Kunming, pp 5–30